

主導課程七：自然語言處理(Natural Language Processing)

課程基本資料

開設學校：清華大學

開授教師：高宏宇

班級人數：__2360__人 (保留 __200__人給開課學校，條件式授權聯盟校不限，封閉式授權每校40人)

開課級別：研究所

授課語言：中文

學分數：3

授權方式：封閉式

上課時間：星期四09:00-12:00

是否接受非同步授課：是

實體期中、末評量時間：12/10 09:00-12:00

遠距上課位置：<https://cool.ntu.edu.tw/courses/41436/modules>

課程網頁：https://github.com/IKMLab/NTHU_Natural_Language_Processing

課程概述

本課程旨在介紹自然語言處理(NLP)和大型語言模型(LLM)的基礎知識和前瞻技術，適合對自然語言技術感興趣的學生。隨著生成式人工智慧技術的快速發展，NLP在各個領域中的應用日益廣泛。

本課程將提供學生NLP理論基礎，並結合實際應用，幫助學生掌握最新的NLP與LLM技術。課程內容主要分為以下幾個部分：

1. 文字處理基礎:介紹NLP的基本概念和常用技術。教學基本的文字處理技術，如分詞、詞性標註、命名實體識別等。
2. 機器學習模型:機器學習基本概念和算法，如線性回歸、決策樹、隨機森林等。介紹如何將機器學習應用於NLP，包括文本分類、情感分析等。
3. 語言模型:語言模型的基本概念與原理，如N-gram模型、Word2Vec等。深度學習在語言模型中的應用，如RNN、LSTM、Transformer等架構。詳細講解BERT、GPT等先進語言模型，並探討其在不同NLP任務中的應用。
4. 前瞻大語言模型技術:介紹大型語言模型的發展歷程與最新研究進展，如GPT-3等。探討這些模型的訓練方法、大規模資料集的使用，以及在不同領域中的能力。介紹輕量化微調技術(PEFT)，如LoRA，並說明其在提高訓練效率和效果方面的優勢。

參考書目

無 (自編講義)

課程內容大綱

週次	日期	課程內容	備註
1	10,Sep.	Introduction to NLP & applications	
2	17,Sep.	Introduction to NLP & applications, word vector	
3	24,Sep.	Introduction to AI & Python - Neural Network, Deep learning	

4	1,Oct.	Introduction to AI & Python - Numpy, Panda, Scikit Learn, PTorch	
5	8,Oct.	NLP with NN - RNN, LSTM	
6	15,Oct.	NLP with NN - Word Embedding	
7	22,Oct.	Introduction to Generative AI	
8	29,Oct.	Introduction to Generative AI - Seq2Seq	
9	5,Nov.	Introduction to Generative AI - BERT, GPT, T5	
10	12,Nov.	Introduction to Generative AI - Transformer	
11	19,Nov.	Large Language Model	
12	26,Nov.	Large Language Model	
13	3,Dec.	Training in PM (pre-trained models) - In context learning	
14	10,Dec.	Training in PM (pre-trained models) - PEFT / LoRA	
15	17,Dec.	Application Implementation & Challenges	
16	24,Dec.	Application Implementation & Challenges	

成績評量方式

- Homework x 5 75%
- Midterm Exam x1 25%

課程要求

建議學生須先具Python程式語言的先備知識。